Typology Characterization of Farmers in Mali

March 9th 2016

Sara Signorelli

Introduction

Africa RISING is testing alternative technology options with heterogeneous populations of farmers that will likely respond to the technologies differently. Creating farm typologies is one approach to design targeted interventions that adequately address the needs of different types of farmers. Notably, creating typologies can help:

- **Identify suitable farms to target innovations (ex-ante)**: we assume that not all innovations are appropriate for all farms, and that structuring into groups would support the identification of technology-specific suitable farming systems.
- **Scale out innovations**: on the basis of the heterogeneity in a population we can formulate extension messages, policies and other incentive schemes to further spread the use of designed innovations.
- **Assess agro-economic effects (ex-post)** Explaining trends and farmer 'behavior' (functional characteristics, including sustainable intensification indicators) and verification of the agro-economic effects of the interventions for different farm types.

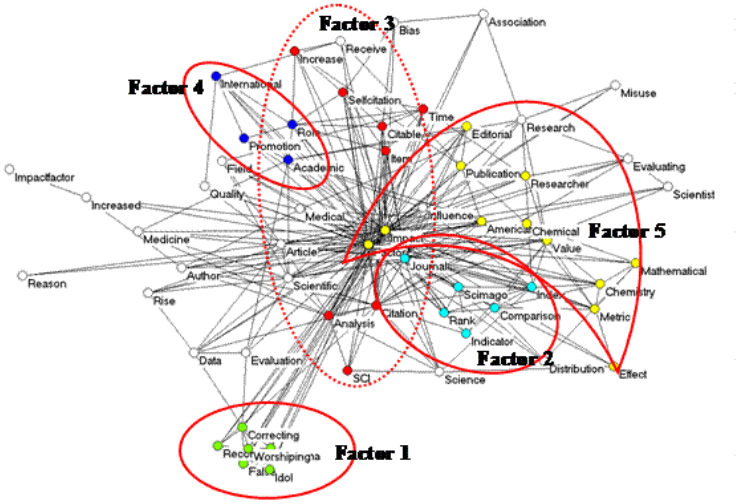
This document presents a summary of a typology study done using quantitative statistical methods (discussed below) applied to micro data from the Mali Africa RISING Baseline Evaluation Survey (MARBES) (conducted in 2014) and secondary data on environmental/biophysical variables from various source. The quantitative approaches have the advantage that they are reproducible and do not impose any ex-ante structure to the clustering process, while more qualitative approaches can potentially incorporate less tangible insights such as cultural patterns. Once the different farm types are identified through systematic quantitative analysis, they need to be validated with input from Africa RISING colleagues (especially working in Mali).

Methodological steps

We apply a combination of factor and cluster analysis to obtain the final groups, or "types" (See Cunningham & Maloney, 1999 for an empirical application). We first use factor analysis to reduce the number of socio-economic variables to characterize the farms by selecting the most relevant ones in differentiating the sample. Factor analysis is often used to discover underlying patterns in data and its aim is to explain the largest portion of the entire dataset variation with the lowest possible number of factors. Factors are unobserved variables that summarize the correlation among several observed variables and factor analysis allows us to divide the dataset into different factors, or dimensions, and categorize each variable into one of the factors. Figure 1 shows an example of how the variables in a dataset are divided into different dimensions to explain the total variation in the data. The analysis

also allows us to rank the factors by their importance in explaining the variation in the data and to further rank each variable by its explanatory power within the factor.

Figure 1: Example of factor analysis[1]



Our factor analysis based on MARBES data involves the following main steps (see for example McDonald; 2014. Basilevsky; 2009. Mulaik; 2009 for a discussion on the methods):

1. We divide the variables in MARBES into the five domains of sustainability that have been identified within Africa RISING to gauge progress: **productivity, economic, environment, social and human**.
2. We perform separate factor analysis on each domain to select the variables that explain the largest portion of the variation in the data.
3. We use scree plots to define the number of factors to look at and, within each of the selected factors, we consider the two variables with the highest absolute values of factor loads, conditional on them being greater than 0.5 (or smaller than -0.5).
4. Finally, we obtain a parsimonious set of socio-economic variables that explain most of the variation in the data and thus are highly relevant in defining the different farm types.

The sub-set of variables obtained using steps (1) to (4) are used to perform a cluster analysis, which divides the total sample into a chosen number of clusters (Kaufman & Rousseeuw; 2009. Romesburg; 2004. Galbraith et Al.; 2002). The numbers of clusters are chosen in order to represent groups that are different enough from each other while ensuring that each group to be included has a sufficient amount of observations. There are several different methods to perform cluster analysis, some hierarchical and some non-hierarchical. We chose the hierarchical method using medians, where the distance between two clusters is calculated as the median distance between all pairs of subjects in the two clusters. The results obtained and the characteristics of each group formed are reported in the next section.

[1] http://www.leydesdorff.net/words/

Results

1. Factor analysis of productivity variables (Sustainability Domain 1)

The scree plot of the factorization of the productivity variables (Figure 2) shows that the first four factors (represented by the first four dots at the top of the line graph) are highly relevant but that the 5th factor starts to be less important in explaining the variation (smaller vertical jump).

Figure 2: Scree plot of productivity variables

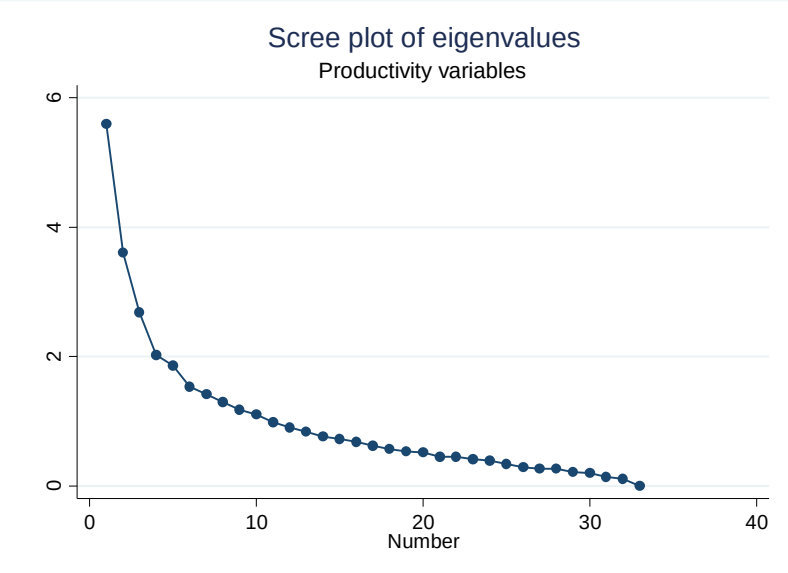


Table 1 shows the rotated matrix of factor loads for the three factors we have chosen, with the relevant variables highlighted (>0.5 or <-0.5). Factor 1 captures elements related to land size, cereal cultivation and fertilizer use. Factor 2 captures elements of crop diversity and intercropping practices. Factor 3 captures livestock and, finally, Factor 4 captures vegetable cultivation and irrigation practices. The final selection of variables for the cluster analysis includes the area cultivated with cereals and the total cereal production in Kg for factor 1, the share of households practicing intercropping and the average number of intercropped plots for factor 2, the share of households breeding mixed livestock and the average number of livestock types owned for factor 3, and finally production and yield of vegetables for factor 4.

Table 1: Factor loads of productivity variables

Variable	Factor1	Factor2	Factor3	Factor 4
Land size (Ha)	0.8409	0.0446	0.1371	-0.0015
N. parcels	0.2888	0.1691	0.2032	0.3458
Min distance plot	-0.123	-0.0116	0.0449	-0.0827
Max distance plot	0.0847	0.0255	0.0345	0.1282
N. trees	0.3378	-0.0062	-0.1001	0.0381
N. crops	-0.0394	0.6733	0.1019	-0.097
N. plots	0.4546	-0.0249	0.1818	0.3512
HH does intercropping	0.0735	0.8535	-0.0139	0.1626
HH does intercropping with legumes	0.0138	0.8342	0.0497	0.0227
N. of intercropped plots	0.0735	0.8535	-0.0139	0.1626
Size intercropped land (Ha)	0.0175	0.651	0.0044	-0.0391
Size legumes-intercropped land (Ha)	-0.0042	0.7448	0.06	-0.1412
Ownership mixed livestock	0.0048	0.033	0.7945	0.0101
N. livestock types owned	0.2211	0.0593	0.852	0.0477
Maize only crop	-0.0594	0.0076	0.1096	-0.0466
Mixed crops	0.0869	0.0297	0.1815	-0.0047
Cultivation of cereals	0.1892	0.0374	0.2545	-0.0545
Cultivation of vegetables	-0.1021	0.0837	0.1548	0.5385
Cultivation of legumes	-0.0026	0.0654	0.1381	0.0708
Area cultivated with cereals (Ha)	0.8423	0.0449	0.0894	-0.0562
Area cultivated with vegetables (Ha)	0.0027	0.0385	0.026	0.0707
Area cultivated with legumes (Ha)	0.3961	0.057	0.0803	-0.0773
Production cereals (Kg)	0.8485	0.0303	0.0836	-0.0041
Production vegetables (Kg)	0.016	0.0778	-0.007	0.7928
Production legumes (Kg)	0.2311	0.0315	0.0076	0.0048
Yield cereals (Kg/Ha)	0.1655	-0.0037	0.0827	0.0595
Yield vegetables (Kg/Ha)	-0.0028	0.0309	0.0044	0.7781
Yield legumes (Kg/Ha)	-0.1102	0.0024	-0.0334	0.1137
TLU small ruminants	0.3485	0.0019	0.6145	-0.0009
TLU big ruminants	0.4625	0.0754	0.4111	0.0689
TLU poultry	0.1105	-0.0365	0.4823	0.0726
Fertilizer used (Kg)	0.7904	0.0172	0.0812	0.0364
HH does irrigation	-0.0338	0.1106	0.0516	0.6354

Note: "N" stands for number. "HH" stands for household. "TLU" stands for Tropical Livestock Units

2. Factor analysis of economic variables (Sustainability Domain 2)

For the economic variables we considered, the relevant factors seem to be the first three (Figure 3). Table 2 shows that factor 1 captures agricultural inputs; factor 2 captures total harvest and its uses; and factor 3 captures non-agricultural wealth and dwelling conditions. The final list of variables considered includes total person-days and male person-days employed in agriculture (factor 1), total harvest quantity and Kg of harvest devoted to own consumption (factor 2), and non-agricultural wealth index and availability of electricity in the dwelling (factor 3).

Figure 3: Scree plot of economic variables

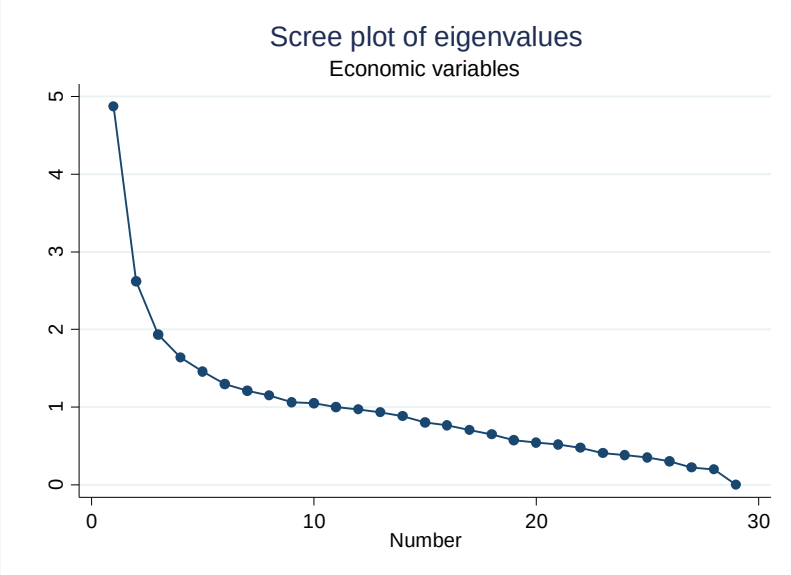


Table 2: Factor loads of economic variables

Variable	Factor1	Factor2	Factor 3
Fertilizer cost	0.5761	0.2502	0.1206
Traditional seeds cost	0.3765	0.1282	0.1481
Improved seeds cost	0.084	0.2062	-0.0367
Pesticide cost	0.3309	0.2674	0.2456
Other non-labor cost	0.0818	0.1185	0.055
Animal feed cost	0.0031	0.1021	0.0881
Agricultural wage	0.0218	-0.0677	0.0369
HH uses community labor	0.0915	-0.045	0.0096
HH uses hired labor	0.0887	0.0424	-0.0237
Total PD used for crops	0.9829	0.0806	0.0063
Male PD used for crops	0.9294	0.1055	0.0177
Female PD used for crops	0.8969	0.0281	-0.0126
Family PD used for livestock	0.0566	-0.0076	0.0891
Hired PD used for livestock	-0.0159	0.087	-0.0036
Total harvest of grains (Kg)	0.1086	0.839	0.0017
Total harvest of stover (Kg)	-0.0079	0.1912	0.0234
Total harvest used for animal feed (Kg)	-0.0208	0.0309	-0.1225
Total harvest used for crop residual (Kg)	-0.0474	-0.047	0.1921
Total harvest used for seeds (Kg)	0.0667	0.6202	0.0097
Total harvest used for gifts (Kg)	0.0953	0.7887	0.0209
Total harvest used for own consumption (Kg)	0.1071	0.8286	0.1112
Total harvest sold (Kg)	-0.0248	0.0474	-0.0508
Total harvest used for other reasons (Kg)	0.1303	0.6796	-0.0066
Agri wealth index	0.2933	0.1471	0.2629
Non-agri wealth index	0.0368	0.0519	0.832
Good floor material in dwelling	-0.0665	0.1249	0.4693
Good source of drinking water	0.0174	-0.0835	0.1019
Good toilet facility	-0.0729	-0.1238	0.0259
Good lightning source	0.0255	0.0286	0.7961

Note: "HH" stands for household and "PD" refers to person-days.

3. Factor analysis of environment variables (Sustainability Domain 3)

For the environment domain, we identified three relevant factors. The first concerns the self-reported experience of the farmers in terms of soil erosion, the second captures the observed characteristics of the parcels and the third measures soil conservation practices as well as tree ownership. Our final selection of variables includes share of households experiencing soil erosion – overall as well as the ones that do not take any mitigating measures - (factor 1), share of parcels with clay or loam soil and with black or brown soil (factor 2), and share of households using crop rotation and total number of trees owned on the land (factor 3).

Figure 4: Scree plot of environment variables

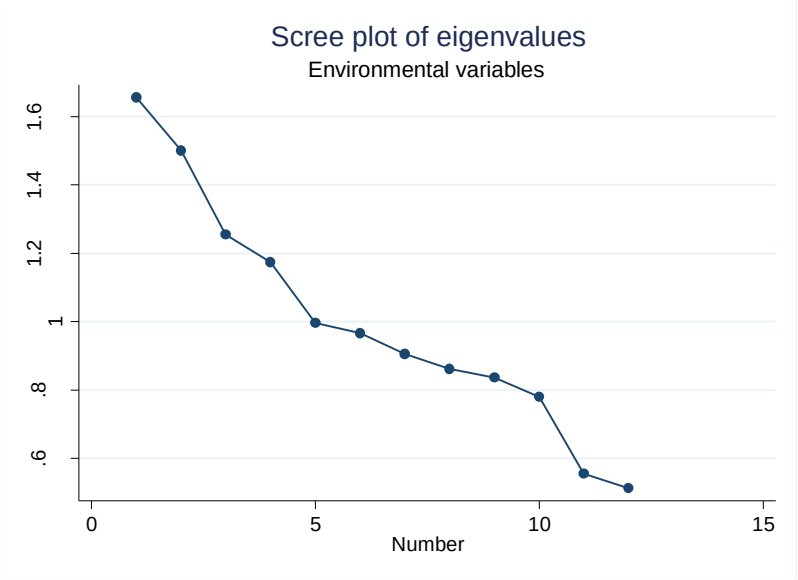


Table 3: Factor loads of environment variables

Variable	Factor 1	Factor 2	Factor 3
HH uses irrigation	0.2986	0.1587	0.3697
HH uses crop rotation	0.0781	-0.0115	0.6841
HH uses fallowing	0.2746	0.0011	-0.0229
HH uses alternative tillage	-0.0977	0.023	0.124
HH uses manure	-0.0394	0.1582	0.4117
HH uses urea	0.1615	-0.1207	0.2161
HH experiences soil erosion	0.8179	-0.0046	0.0528
HH experiences soil erosion and does not takes any preventive measure	0.7428	-0.0719	-0.0053
Share of parcels with clay or loam soil	-0.0005	0.839	0.0135
Share of parcels with black or brown soil	-0.0456	0.8305	-0.0003
Share of parcels with incrusted soil	0.4158	0.0232	-0.2796
Number of trees owned	-0.0715	-0.023	0.5843

4. Factor analysis of social variables (Sustainability Domain 4)

Our dataset has a relatively small set of variables capturing social aspects, focusing on gender disparities. We thus chose only the first factor, which highlights the presence of females and females-only managed plots as the main variables of interest.

Figure 5: Scree plot of social variables

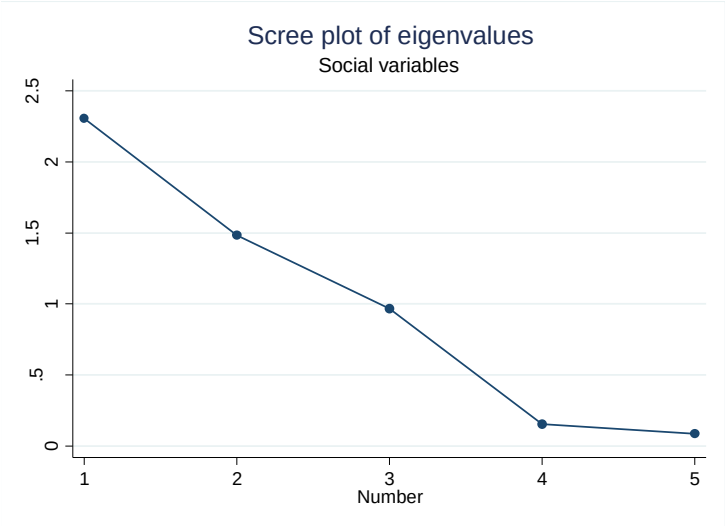


Table 4: Factor loads of social variables

Variable	Factor1
Females also responsible for plots	0.9643
Females only responsible for plots	0.9632
Females also responsible for livestock	0.1208
Females only responsible for livestock	0.1024
Wage gap	-0.1844
Part of compound (proxy by farming communal land)	-0.1059

5. Factor analysis of human variables (Sustainability Domain 5)

The final sustainability domain we focus on is human capital. We select the first five factors, which capture the age composition of household members, including the prevalence of younger age groups from 0 to 29 years old (factor 1) and older age groups above 45 years old (factor 2), the level of education of household members (factor 3), the characteristics of the household head (factor 4) and the size of the household (factor 5). Experiencing food shortages in the 12 months preceding interview date do not appear to play a key role in differentiating the sample. We finally select young and total dependency ratio (factor 1), mean adult age and share of over 45 in the household (factor 2), and mean level of education in the household and years of education of the household head (factor 3), the

indicator of whether the household head is single and whether he's male and single (factor 4) and the size of the household and number of male adults (factor 5).

Figure 6: Scree plot of human variables

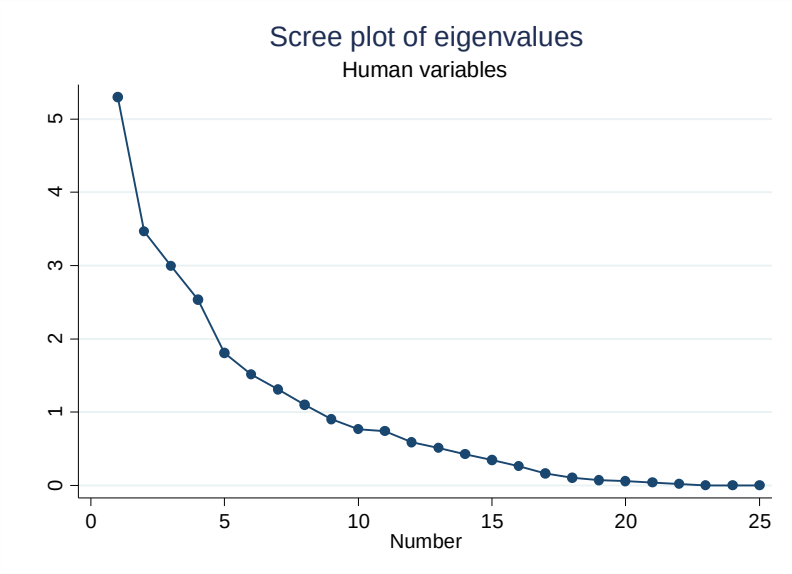


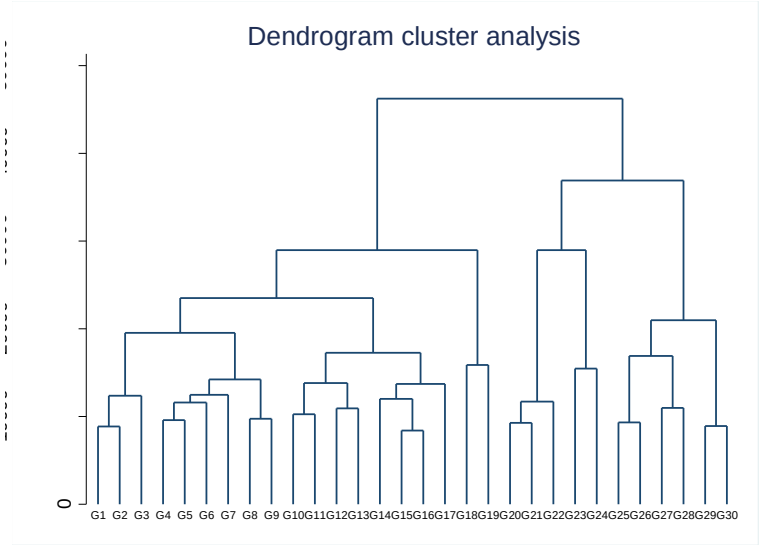
Table 5: Factor loads of human variables

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
HH size	0.5685	-0.0614	0.1147	-0.0659	0.7819
Head is married	-0.0325	0.0732	0.0013	-0.6519	0.1102
Head is widow	0.0069	0.0132	0.0337	0.0154	0.0033
Head is single	-0.0559	-0.0071	0.045	0.9782	-0.0181
Head is female	-0.0531	0.0606	0.0209	-0.0482	-0.0373
Head is male and single	-0.0559	-0.0071	0.045	0.9782	-0.0181
Head's age	-0.0512	0.8103	0.0994	-0.1433	0.4232
Head's years of educ	0.0099	-0.0467	0.8992	0.0331	-0.1269
Head is literate	0.1161	-0.0408	0.4454	0.0574	-0.0894
Mean years of edu.	-0.1457	-0.0235	0.9127	0.0787	0.1448
Highest years of edu.	-0.1769	0.0445	0.8496	-0.0043	0.322
Mean age	-0.5255	0.8162	-0.0062	0.0393	-0.0629
Mean adult age	0.1608	0.9448	-0.0704	0.0132	-0.0968
N. of males adults	-0.2844	0.0875	0.1307	0.0196	0.8068
N. of females adults	-0.0577	0.0355	0.178	-0.1395	0.5957
children	0.6925	-0.3467	-0.0746	-0.0621	0.0965
Young dep. Ratio	0.9634	-0.0845	-0.0799	-0.0193	-0.0548
Old dep. Ratio	0.06	0.559	-0.0072	0.0339	-0.092
Total dep. ratio	0.9631	0.0139	-0.0803	-0.0132	-0.0702
Share of 0-14 y.o.	0.9257	-0.2166	-0.067	-0.0571	0.0691
Share of 15-29 y.o.	-0.6777	-0.4816	0.126	0.0159	-0.0074
Share of 30-44 y.o.	0.0934	-0.1397	-0.0789	0.0743	-0.117
Share of > 45 y.o.	-0.305	0.863	-0.0137	-0.0076	0.0148
HH worries for food shortages	-0.0076	-0.0137	0.0568	0.0275	0.0299
Months experienced food shortages	0.0502	0.1092	-0.0211	0.0243	-0.0096

6. Cluster analysis

The analysis summarized in the preceding section informed the selection of a list of factors that we used in the cluster analysis. These are 8 productivity variables, 6 economic variables, 6 environmental variables, 2 social variables and 10 human variables. Figure 7 shows the dendrogram illustrating how the farm households in our sample can be split into different groups (or types) based on these variables we have identified. The vertical distance between separations illustrates the distance of the different groups to each other.

Figure 7: Dendrogram



Considering the number of observations within each group and differentiation of characteristics between groups, we decided to create four final groups, or "types" of farmers. Tables 6a-6e illustrate the distribution of characteristics across these types and sustainability domains discussed before. Because the clusters were defined using the variables accounting for most of the data variation, as captured by the factor analysis, most of the characteristics differ significantly across every type. Type 1 is the biggest one and includes 261 of the farmers in the sample (37%). Type two defines 171 farmers (24%), type 3 accounts for 135 farmers (19%) and, finally, type 4 defines 138 farmers (20%).

Table 6a: distribution of characteristics in the economic domain

	Type 1	Type 2	Type 3	Type 4
Economic Domain				
Value of fertilizer used (CFAF)	96465.11***	214442.76	327636.00***	375360.79***
	[5672.68]	[10972.63]	[17285.65]	[20283.75]
Value of traditional seeds purchased (CFAF)	2096.16***	3121.67	5107.85**	6291.89***
	[380.24]	[517.23]	[787.68]	[886.53]
Value of improved seed purchased (CFAF)	637.74***	1965.8	3693.07***	3195.26**
	[166.84]	[398.50]	[762.39]	[654.35]
Value of pesticides used (CFAF)	13405.41***	23967.46	36201.48***	47525.44***
	[1395.62]	[2480.36]	[3587.63]	[3984.31]
Share of households using communal labor	0.41**	0.49	0.4	0.59***
	[0.03]	[0.04]	[0.04]	[0.04]
Share of households using hired labor	0.29***	0.37	0.39	0.44**
	[0.03]	[0.04]	[0.04]	[0.04]
Total person-days used, male & female	268.56***	464.96	571.06***	787.29***
	[13.43]	[24.34]	[30.51]	[58.41]
Total Kg of grains harvested	1976.41***	1777.99***	10497.65***	2154.35***
	[121.93]	[152.71]	[388.67]	[205.40]
Total Kg harvest used for own consumption	867.93***	985.30***	4603.56***	1462.92**
	[66.14]	[137.71]	[321.19]	[258.17]
Total Kg harvest sold	718.62***	997.98***	3418.24***	1708.58
	[125.46]	[134.85]	[296.38]	[293.66]
Agricultural wealth index	-0.51***	-0.1	0.50***	0.54***
	[0.04]	[0.06]	[0.09]	[0.11]
Non-agricultural wealth index	-0.21***	-0.08	0.31***	0.20***
	[0.06]	[0.07]	[0.10]	[0.09]
Share of households with good floor in dwelling	0.13	0.09**	0.21***	0.14
	[0.02]	[0.02]	[0.04]	[0.03]
Share of households with good source of drinking water	0.02	0.03	0.01	0.06**
	[0.01]	[0.01]	[0.01]	[0.02]
Share of households with good toilet facility	0.06	0.04	0.05	0.04
	[0.01]	[0.02]	[0.02]	[0.02]
Share of households with good source of lighting	0.35*	0.34*	0.46*	0.48**
	[0.03]	[0.04]	[0.04]	[0.04]
N. of observations	261	171	135	138

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6b: distribution of characteristics in the productivity domain

	Type 1	Type 2	Type 3	Type 4
Productivity Domain				
Total land size (Ha)	5.27***	9.73	14.06***	15.93***
	[0.27]	[0.32]	[0.60]	[0.62]
Share of households doing intercropping	0.02***	0.05	0.03	0.12***
	[0.01]	[0.02]	[0.01]	[0.03]
Share of households doing intercropping with legumes	0.01	0.02	0.01	0.04**
	[0.01]	[0.01]	[0.01]	[0.02]
Area of intercropped plots	0.03**	0.22	0.11	0.51***
	[0.02]	[0.12]	[0.07]	[0.24]
Area of plots intercropped with legumes	0.01	0.05	0.04	0.05
	[0.00]	[0.05]	[0.03]	[0.03]
Share of households owning mixed livestock	0.81***	0.89	0.96***	0.91
	[0.02]	[0.02]	[0.02]	[0.02]
N. of different livestock types owned	2.83***	3.46	3.96***	3.78***
	[0.09]	[0.10]	[0.11]	[0.12]
Share of households cultivating maize only	0.01	0.01	0	0
	[0.01]	[0.01]	[0.00]	[0.00]
Share of households growing cereals	0.93***	1.00**	1.00**	1.00**
	[0.02]	[0.00]	[0.00]	[0.00]
Share of households growing vegetables	0.38	0.39	0.42	0.4
	[0.03]	[0.04]	[0.04]	[0.04]
Share of households growing legumes	0.76***	0.84	0.87*	0.83
	[0.03]	[0.03]	[0.03]	[0.03]
Area of cereals(ha)	3.33***	6.03*	8.62***	11.54***
	[0.28]	[0.23]	[0.46]	[0.54]
Area of vegetables(ha)	0.27	0.26	0.28	0.42
	[0.11]	[0.08]	[0.09]	[0.22]
Area of legumes(ha)	1.28***	1.64	2.65***	2.70***
	[0.18]	[0.16]	[0.37]	[0.29]
Production of cereals(kg)	1878.90***	5993.36**	9574.07***	14437.57***
	[76.12]	[111.23]	[460.90]	[547.40]
Production of vegetables(kg)	80.85	124.33	94.54	117.64
	[13.73]	[38.50]	[33.22]	[27.50]
Production of legumes(kg)	660.72***	1039.15	1463.26***	1311.47***
	[54.27]	[91.69]	[143.64]	[124.58]
Yield of cereals(kg/ha)	809.20***	1172.73	1282.50***	1416.85***
	[35.66]	[39.24]	[46.31]	[45.39]
Yield of vegetables(kg/ha)	994.07	1025.07	1160.64	1142.51
	[206.96]	[240.07]	[318.40]	[264.14]
Yield of legumes(kg/ha)	816.59	897.38	919.54	790.52
	[52.61]	[64.84]	[67.64]	[64.52]
TLU small ruminants	0.42***	0.63*	1.15***	1.13***
	[0.04]	[0.06]	[0.12]	[0.11]
TLU big ruminants	2.11***	3.53**	7.87***	8.10***
	[0.25]	[0.30]	[0.87]	[0.80]
TLU poultry	0.10***	0.13	0.17**	0.19***
	[0.01]	[0.01]	[0.02]	[0.03]
Kg fertilizer used	354.51***	848.92	1235.90***	1474.06***
	[21.64]	[40.94]	[63.78]	[78.84]
N. of observations	261	171	135	138

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6c: distribution of characteristics in the social domain

	Type 1	Type 2	Type 3	Type 4
Social Domain				
Share of HH with female having shared plot responsibility	0.28***	0.18**	0.17*	0.15***
	[0.02]	[0.02]	[0.02]	[0.02]
Share of HH with female having exclusive plot responsibility	0.26***	0.15***	0.17	0.14***
	[0.02]	[0.01]	[0.02]	[0.01]
Share of HH with female having shared livestock responsibility	0.10**	0.06	0.06	0.06
	[0.01]	[0.01]	[0.01]	[0.01]
Share of HH with female having exclusive livestock responsibility	0.07*	0.04	0.03	0.05
	[0.01]	[0.01]	[0.01]	[0.01]
*Gender wage gap (wage women/wage men*100)*	73.4	62.43	77.16	77.25
	[7.57]	[9.27]	[4.01]	[7.34]
Household is member of a compound (farms land belonging to the compound)	0.44***	0.63	0.83***	0.82***
	[0.03]	[0.04]	[0.03]	[0.04]
N. of observations	261	171	135	138

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6d: distribution of characteristics in the human domain

	Type 1	Type 2	Type 3	Type 4
Human Domain				
Household size	5.54	5.54	5.64	5.78
	[0.15]	[0.17]	[0.17]	[0.19]
Share of married heads	0.92***	0.98	1.00**	1.00**
	[0.02]	[0.01]	[0.00]	[0.00]
Share of female heads	0.03***	0.01	0.01	0.00*
	[0.01]	[0.01]	[0.01]	[0.00]
Age of the head	43.54	44.69	44.9	43.29
	[0.97]	[1.07]	[1.17]	[0.97]
Years of education of the heads	2	2.37	2.49	2.81
	[0.25]	[0.36]	[0.45]	[0.41]
Share of literate heads	0.32*	0.37	0.41	0.41
	[0.03]	[0.04]	[0.04]	[0.04]
Mean years of education in the household	2.35	2.64	2.2	2.76
	[0.18]	[0.23]	[0.24]	[0.26]
Max years of education in the household	4.95	5.35	4.74	5.22
	[0.33]	[0.44]	[0.48]	[0.45]
Average age of adults in the household	34.7	33.97	34.73	33.74
	[0.60]	[0.65]	[0.69]	[0.75]
Number of children in the household	1.17	1.19	1.23	1.32
	[0.06]	[0.07]	[0.09]	[0.09]
Young dependency ratio	1.2	1.14	1.22	1.22
	[0.05]	[0.06]	[0.08]	[0.07]
Old dependency ratio	0.03	0.05*	0.05	0.01**
	[0.01]	[0.01]	[0.01]	[0.01]
Share of HH worrying about food shortages	0.22***	0.09**	0.14	0.09**
	[0.03]	[0.02]	[0.03]	[0.02]
Months experiencing food shortages?	0.64***	0.30*	0.41	0.25**
	[0.10]	[0.07]	[0.08]	[0.06]
N. of observations	261	171	135	138

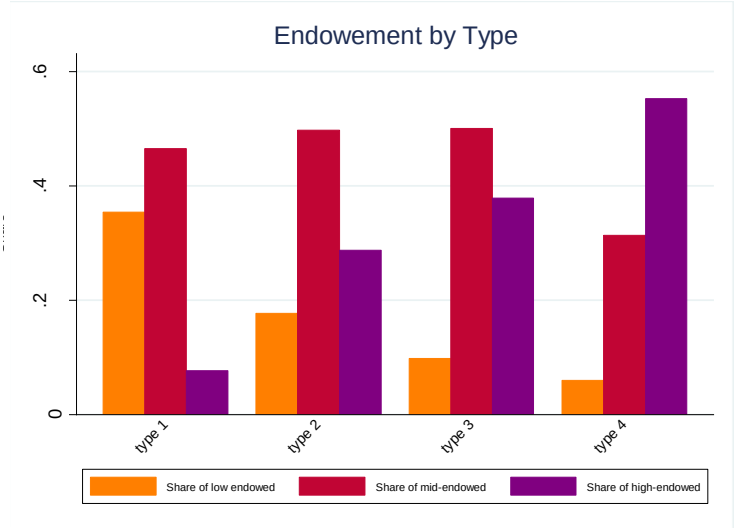
Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6e: distribution of characteristics in the environment domain

	Type 1	Type 2	Type 3	Type 4
Environmental Domain				
Share of households practicing irrigation	0.11	0.08	0.1	0.12
	[0.02]	[0.02]	[0.03]	[0.03]
Share of households practicing rotation	0.80***	0.92	0.96***	0.97***
	[0.02]	[0.02]	[0.02]	[0.01]
Share of households practicing fallowing	0.12	0.08	0.1	0.07
	[0.02]	[0.02]	[0.03]	[0.02]
Share of households practicing alternative tillage	0.06	0.05	0.01*	0.05
	[0.01]	[0.02]	[0.01]	[0.02]
Share of households using manure on (any) plot in either season	0.51***	0.78***	0.80***	0.80***
	[0.03]	[0.03]	[0.03]	[0.03]
Share of households using urea on (any) plot in either season	0.28***	0.37	0.36	0.44**
	[0.03]	[0.04]	[0.04]	[0.04]
Share of households affected by soil erosion	0.28	0.26	0.36*	0.35
	[0.03]	[0.03]	[0.04]	[0.04]
Share of households with soil erosion but no erosion control measu	0.08	0.06	0.07	0.09
	[0.02]	[0.02]	[0.02]	[0.02]
Average share of parcels with clay or loam soil	0.47	0.52	0.44	0.51
	[0.02]	[0.03]	[0.03]	[0.03]
Average share of parcels with black or brown soil	0.68*	0.77**	0.74	0.7
	[0.02]	[0.03]	[0.03]	[0.03]
Average share of parcels with incrusted soil	0.28***	0.16**	0.22	0.14***
	[0.02]	[0.02]	[0.03]	[0.02]
Number of trees owned on the land	64.40***	108.99	113.35	137.27***
	[6.56]	[13.51]	[11.47]	[12.38]
N. of observations	261	171	135	138

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 8: Level of Endowments by Type



The four types differ from each other across all of the five domains, as shown in table 6. One of the striking characteristic that stands out in differentiating them is the level of endowments, as measured by a wealth index including dwelling characteristics, size of the cultivated land and ownership of

agricultural and non-agricultural assets (figure 8). We defined low-endowed households as the ones in the bottom quartile of the wealth distribution, mid-endowed households as the ones in the 2nd and 3rd quartile and highly endowed households as the ones in the top quartile of the asset distribution. Figure 8 shows in which of the endowments category fall most of the households in our typologies. More broadly, the types can be characterized as following:

Type 1: Low-endowment households with low agricultural production

- Slightly higher share of female-headed households (even if still only 3%). Low literacy rates and high food insecurity.
- Higher proportion of women with plot and livestock responsibilities. Smaller proportion of households that are members of a larger compound.
- Little asset ownership: small land holdings, little livestock, low agricultural and non-agricultural wealth.
- Low production of all major crops, also due to low input use (both in terms of labor inputs and non-labor inputs).
- High percentage of incrusted soils, little use of urea and manure.

Type 2: Mid-endowment households growing vegetables and doing intercropping

- High gender inequality, especially in terms of wage gaps.
- Frequent intercropping practices and high share of vegetables growers.
- High vegetable production and productivity.
- Medium level of endowments.
- Medium input use but frequent employment of hired and community labor.
- Little use of soil conservation practices but also little problems of soil quality.

Type 3: High productivity households with high levels of non-agricultural wealth

- Households with fairly low levels of education.
- Higher proportion of households that are members of a larger compound with respect to type 1 and 2.
- Large production and productivity, especially of legumes, but seldom use of intercropping.
- Breeders of different types of livestock.
- High levels of wealth, especially non-agricultural, and high commercialization of agricultural production.
- Relatively high levels of soil erosion and incrustation, but also use of conservation practices to undermine the problem.

Type 4: Highly endowed households producing cereals and breeding livestock

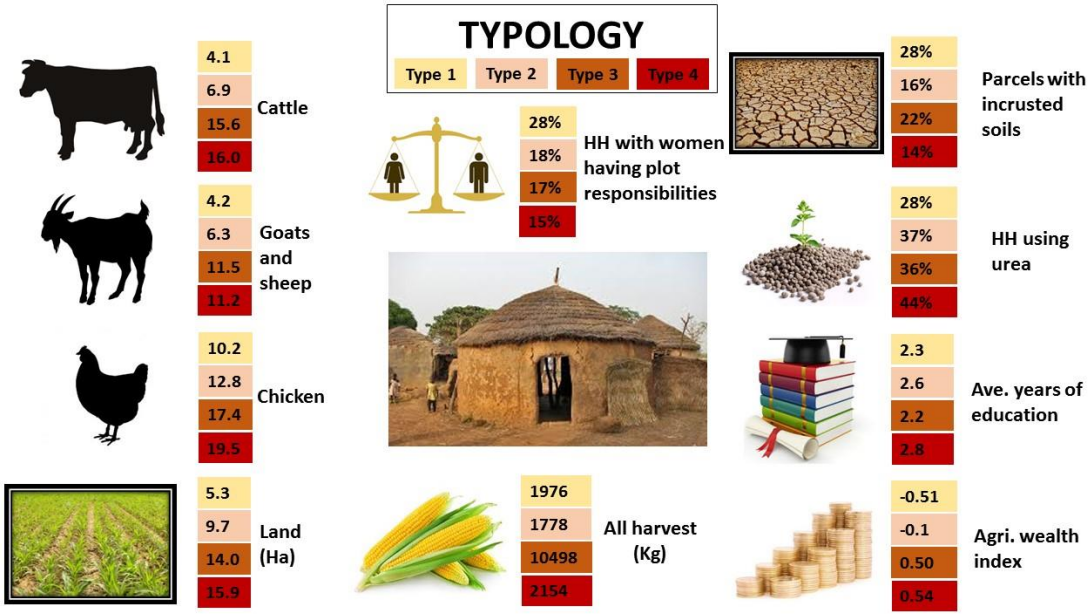
- Young, well-educated, food secure households with many children.
- Higher proportion of households that are members of a larger compound with respect to type 1 and 2.
- Low levels of gender equality across all measurements.
- Largest land holdings and numerous livestock units of both large and small ruminants and poultry.
- High production of crops, especially cereals, with high input use.
- Very high levels of agricultural wealth, fairly high levels of non-agricultural wealth.
- Relatively high levels of soil erosion and lack of mitigating measures.

Table 7 summarizes the main characteristics of every type relative to each sustainability domain, providing a simplified framework for classifying farm households into a particular type. Figure 9 shows a graphic representation of the main characteristics of each type.

Table 7: matrix of performance for each SI domain

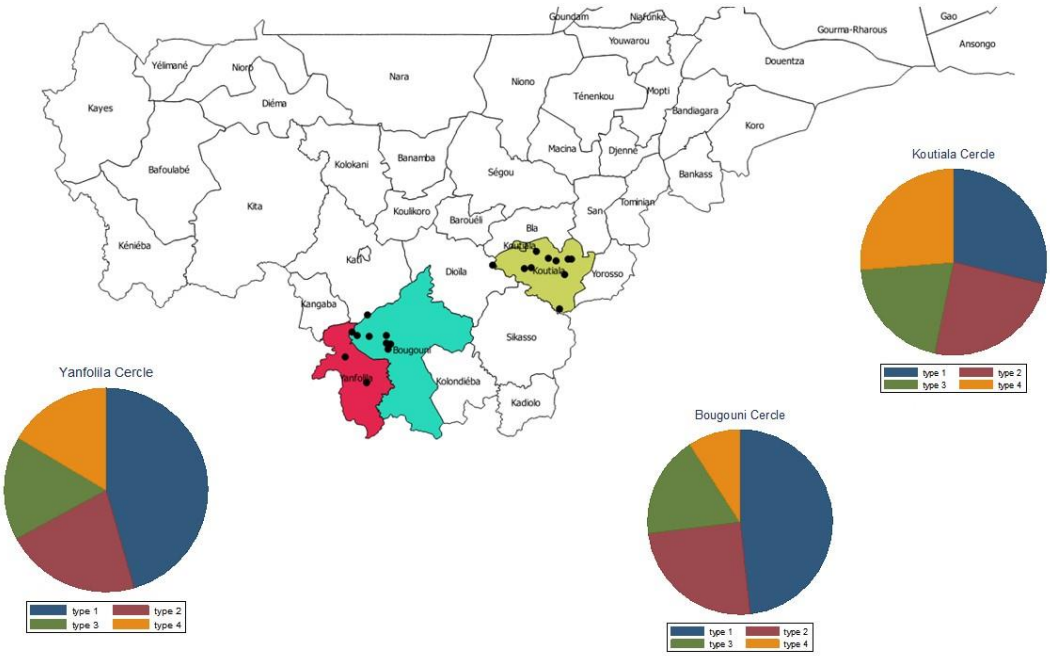
	Productivity	Economic	Environment	Social (gender)	Human
Type 1: Low-endowment households with low agricultural production	Low crop production. Little livestock owned.	Low wealth (agri and non-agri), low input expenditure.	Incrusted soils and little use of soil conservation practices.	High frequency of female responsibility for resources. Less likely to be part of a compound.	Low literacy rates. Slightly higher share of female heads. High food insecurity.
Type 2: Mid-endowment households growing vegetables and using intercropping	High production and productivity of vegetables, little production of cereals and legumes. Frequent use of intercropping.	Medium wealth (agri and non-agri), low input expenditure but higher than type 1. Frequent use of hired and community labor.	Little use of soil conservation practices but also fairly good soil quality.	Low levels of gender equality.	Fairly high education levels and relatively low food insecurity.
Type 3: High productivity households with high levels of non-agricultural wealth	Large production and productivity of legumes. Breeding of different livestock. Little intercropping.	High wealth, especially non-agric., large total production and commercialization of crops.	Relatively high levels of soil incrustation but frequent use of conservation practices.	Average gender equality. More likely to be part of a larger compound.	Fairly low levels of education.
Type 4: Highly endowed households producing cereals and breeding livestock	Largest land and livestock holdings. Very large production of all main crops but especially cereals.	High agri. wealth and fairly high non-agri. wealth. High input use (especially fertilizer).	High frequency of soil erosion for which no measures are taken but frequent use of conservation practices.	Low levels of gender equality. More likely to be part of a larger compound.	Young households with many children and high levels of education. Low food insecurity.

Figure 9: Graphic representation of types



The different types are not homogeneously distributed across space. Figure 10 shows the typology composition of each cercle in the sample. While in Bougoni there is a high concentration of households with low levels of endowments (type 1 and 2), Koutiala concentrates high shares of Mid-endowed and high endowed households (type 3 and 4). The spatial distinctions are important because they can support interventions based on the most prevalent households' typologies in the area.

Figure 10: Distribution of Typologies by cercles



The characteristics of each household type described above can be displayed clearly with a spider plot. Figure 11 summarizes the performance of each type relative by each domain as follows:

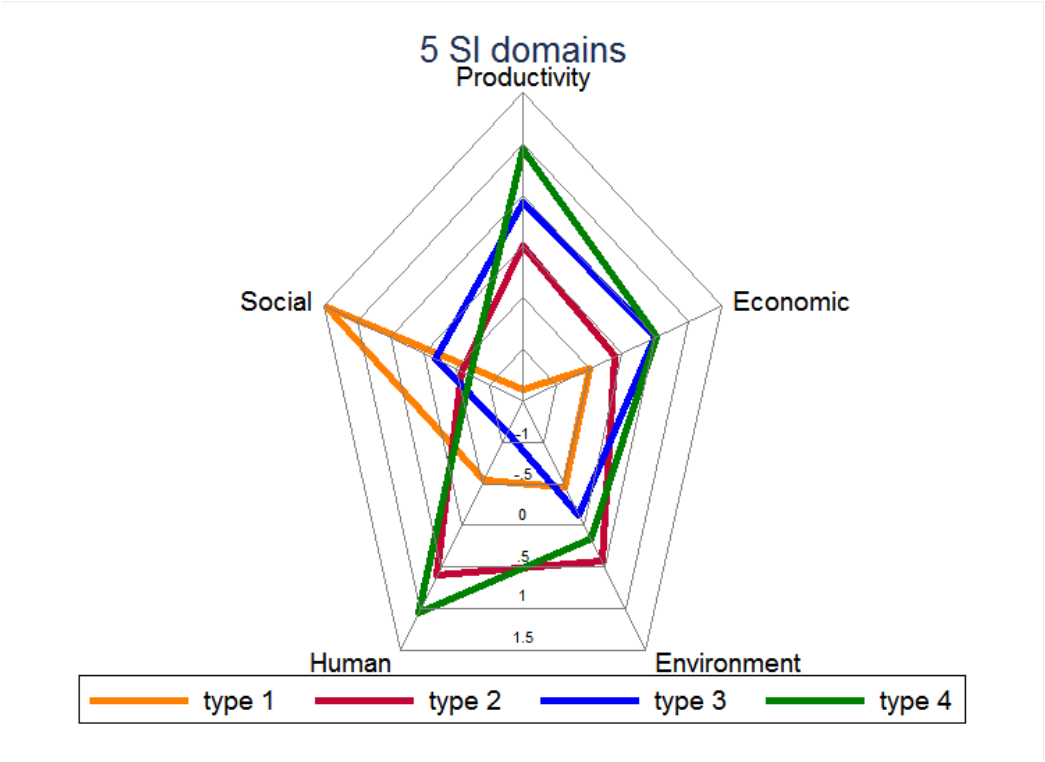
- Type 3 and 4 are the best performers in terms of economics and productivity, but type 3 lags behind in terms of environmental conditions and human endowments while type 4 performs the worst in gender equality.
- Type 1 includes the largest portion of the farmers in the sample, which lack of most of the endowments considered for sustainable intensification. The social aspect makes exception since type 1 reports the highest scores, as measured by the gender equality indicators. However, it is likely that the high level of women's responsibility in this group is more driven by necessity than by choice.
- Finally, type 2 presents favorable environmental conditions and high levels of human endowments, but cannot capitalize on these strengths to achieve high productivity and economic performance.

Recommendations:

- The "low-hanging fruits" emerging from this typology study can be found in type 2 and 3. Type 2 is well educated and possesses land with good quality soil. This group has the potential to rapidly achieve higher productivity and economic outcomes if productive resources are made available to these farmers. AR can focus on introducing new agricultural technologies aiming at increasing the agricultural production of this group. Type 3, on the other hand, is already on a good track in terms of productive capacity and economic endowments but suffers from bad quality soils, which can potentially reduce his productivity on the long term, and from low levels of human capital. In this case AR can focus on delivering trainings on soil conservation, health and nutrition.
- Type 4 includes farmers that are performing well across most of the SI domains. AR in this case can focus on delivering trainings on the importance of soil conservation and could explore the possibility to involve these farmers in the technology diffusion process by making them trainers of other farmers. Since they are the most successful they are also the best suited to show the example to others and generate higher adoption of the technologies showcased.
- Finally, type 1 includes the majority of AR farmers and requires and integrated approach including both the introduction of new agricultural technologies, trainings on soil conservation, health and nutrition, and the establishment of a support system that can quickly respond to the farmers' needs.
- The gender equality aspect seems to be particularly delicate in the case of Mali. In order to improve women's access to resources AR should consult closely with the gender experts of the program, who should be directly involved in the design of interventions targeting women's inclusion directly.

The appendix includes additional graphs characterizing the obtained typologies.

Figure 10: Typologies performance by sustainability domain



NOTE: The following variables are used to measure each domain: cereals yield (Productivity), asset-based wealth index (Economic), soil conservation index composed of crop rotation, alternative tillage, experience of soil erosion without measures for mitigating it and share of parcels with incrusted soils (Environment); gender equality index composed by female responsibility in managing certain plots and livestock (Social), and average education in the household (Human).

References

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications* (Vol. 418). John Wiley & Sons.

Cunningham, W., & Maloney, W. F. (1999). Heterogeneity among Mexico's micro-enterprises: an application of factor and cluster analysis. *World Bank Policy Research Working Paper*, (1999).

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863-14868.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. CRC Press.

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago Press.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues* (Vol. 14). Sage.

McDonald, R. P. (2014). *Factor analysis and related methods*. Psychology Press.

Mulaik, S. A. (2009). *Foundations of factor analysis*. CRC press.

Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu.com.

Sethi, S. P. (1971). Comparative cluster analysis for world markets. *Journal of Marketing Research*, 348-354.

Appendix Figures

Figure A1: Typologies by domain (productivity and economic)

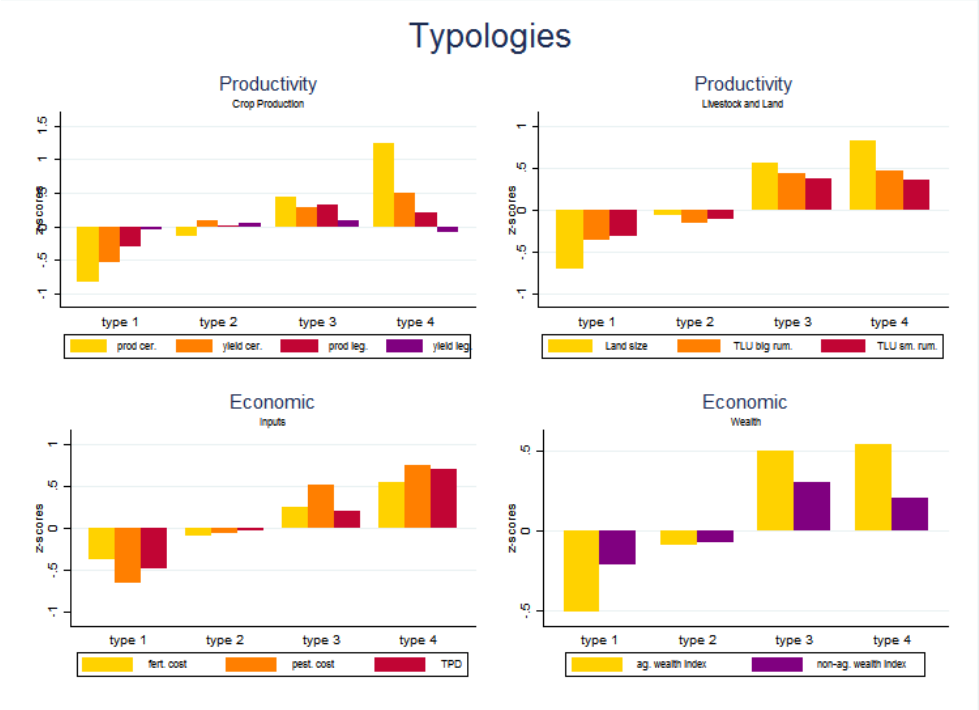


Figure A2: Typologies by domain (environment, social and human)

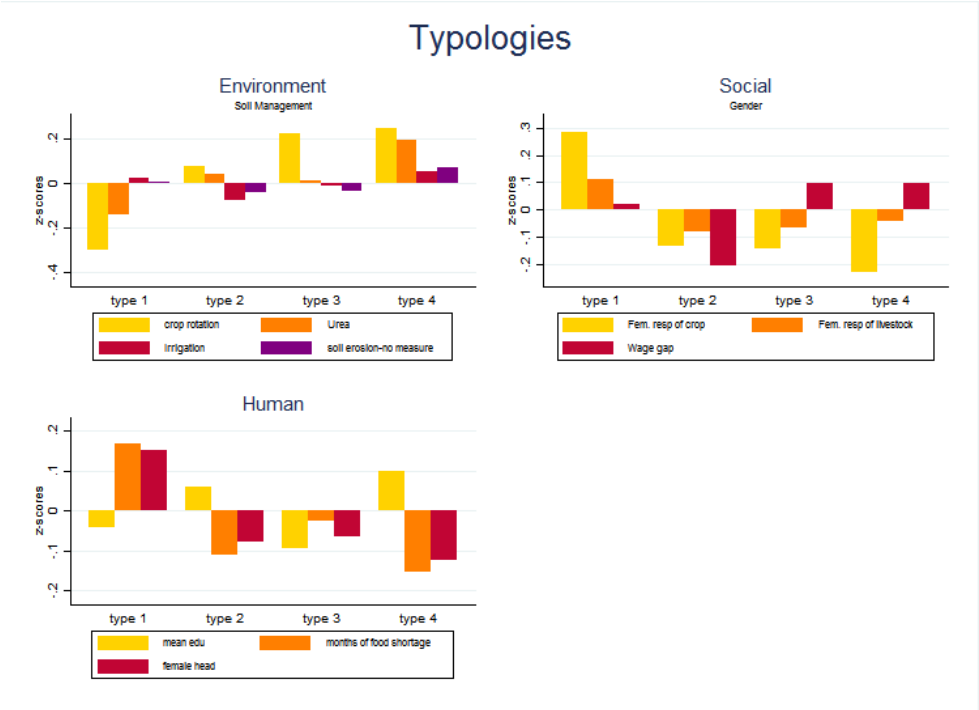


Figure A3: Radar graph – productivity (z-scores)

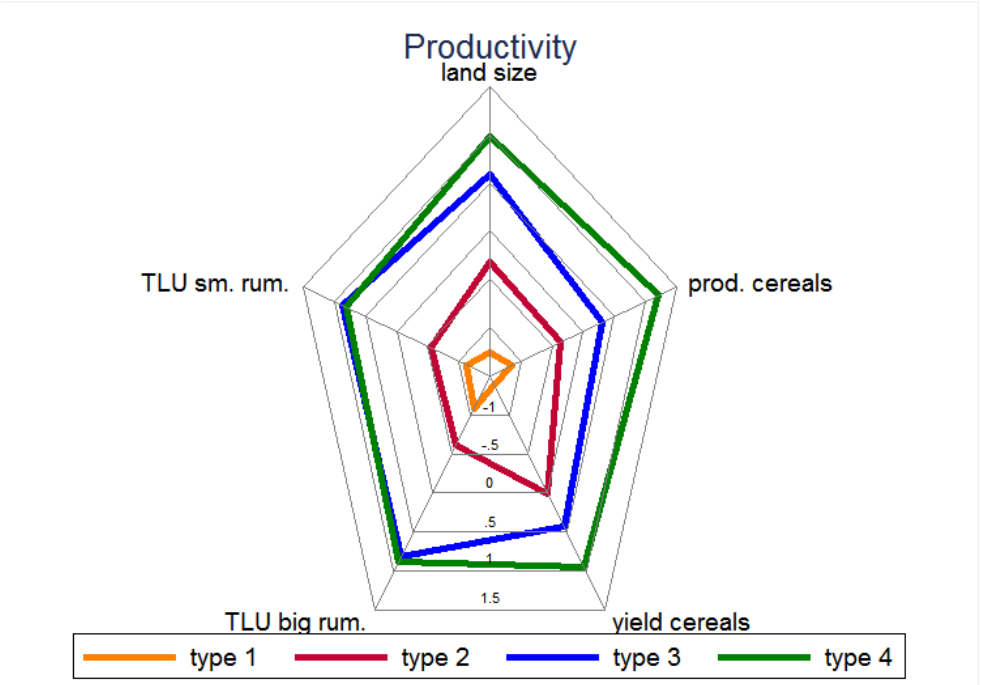


Figure A4: Radar graph – economic (z-scores)

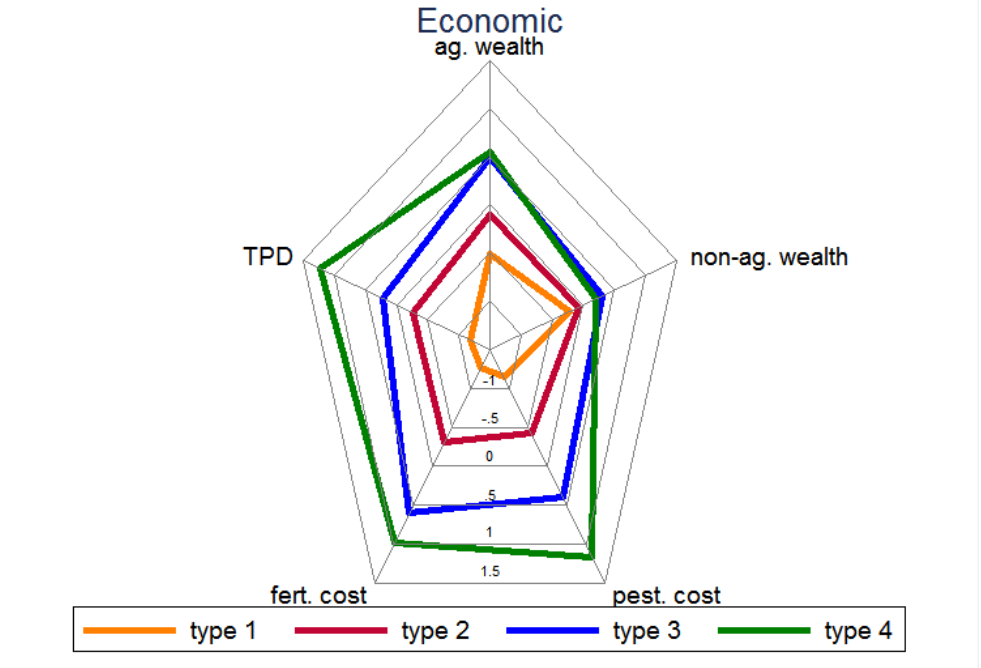


Figure A5: Radar graph – environment (z-scores)

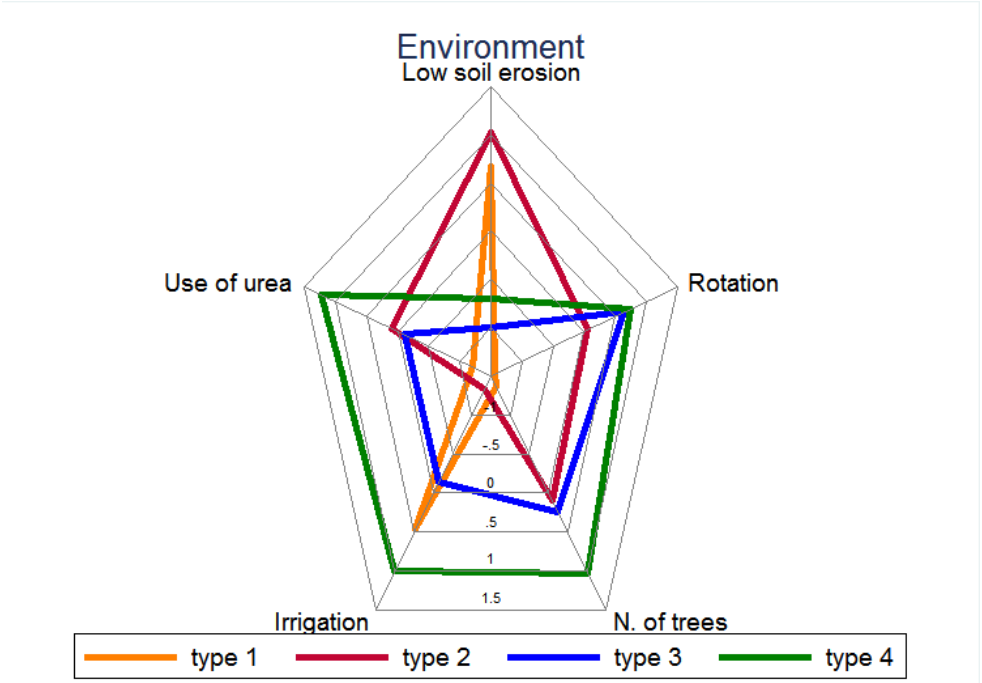


Figure A6: Radar graph – social and human (z-scores)